working paper
department
of economics

*INFORMATIONAL HERDING*
*AND*
*OPTIMAL EXPERIMENTATION*

Lones Smith
Peter Sorensen

No. 97-22 October , 1997

massachusetts
institute of
technology

50 memorial drive
cambridge, mass. 02139

WORKING PAPER
DEPARTMENT
OF ECONOMICS

# *INFORMATIONAL HERDING AND OPTIMAL EXPERIMENTATION*

Lones Smith
Peter Sorensen

No. 97-22 October , 1997

MASSACHUSETTS
INSTITUTE OF
TECHNOLOGY

50 MEMORIAL DRIVE
CAMBRIDGE, MASS. 02139

# *Informational Herding and Optimal Experimentation*[*]

Lones Smith[†]
Economics Dept., M.I.T

Peter Sørensen[‡]
Nuffield College, Oxford

October 30, 1997

**Abstract**

We explore the constrained efficient observational learning model — as when individuals care about successors, or are so induced by an informationally-constrained social planner. We find that when the herding externality is correctly internalized in this fashion, incorrect herds still obtain.

To describe behaviour in this model, we exhibit a set of indices that capture the privately estimated social value of every action. The optimal decision rule is simply: Choose the action with the highest index. While they have the flavour of Gittins indices, they also incorporate the potential to signal to successors. We then apply these indices to establish a key comparative static, that the set of stationary 'cascade' beliefs strictly shrinks as the planner grows more patient. We also show how these indices yield a set of history-dependent transfer payments that decentralize the constrained social optimum.

The lead inspiration for the paper is our proof that informational herding is but a special case of myopic single person experimentation. In other words, the incorrect herding outcome is not a new phenomenon, but rather just the familiar failure of complete learning in an optimal experimentation problem.

[*]A briefer paper without the index rules and associated results appeared under the title "Informational Herding as Experimentation Déjà Vu." Still earlier, the proposed mapping appeared in July 1995 as a later section in our companion paper, "Pathological Outcomes of Observational Learning." We thank Abhijit Banerjee, Meg Meyer, Christopher Wallace, and seminar participants at the MIT theory lunch, the Stockholm School of Economics, the Stony Brook Summer Festival on Game Theory (1996), Copenhagen University, and the 1997 European Winter Meeting of the Econometric Society (Lisbon) for comments on various versions. Smith gratefully acknowledges financial support for this work from NSF grants SBR-9422988 and SBR-9711885, and Sørensen equally thanks the Danish Social Sciences Research Council.

[†]e-mail address: `lones@lones.mit.edu`

[‡]e-mail address: `peter.sorensen@nuf.ox.ac.uk`

# 1. INTRODUCTION

The last few years has seen a flood of research on informational herding. We ourselves have actively participated in this research herd (Smith and Sørensen (1997a), or SS) that was sparked independently by Banerjee (1992) and BHW: Bikhchandani, Hirshleifer, and Welch (1992). The context is seductively simple: An infinite sequence of individuals must decide on an action choice from a finite menu. Everyone has identical preferences and menus, and each may condition his decision both on his endowed private signal about the state of the world, and on all predecessors' decisions (but cannot see their private signals).

In this context, SS showed that beliefs converge upon a *cascade* — i.e. where only one action is taken with probability one.[1] BHW and Banerjee showed that a 'herd' eventually arises — namely, after some point, all decision-makers ($\mathcal{DM}$s) make the identical choice, possibly unwise. Clarifying their next result, SS also showed that this herd is ex post incorrect with positive probability iff the $\mathcal{DM}$s' private signals are uniformly bounded in strength. This simple pathological outcome has understandably attracted much fanfare.

The main thrust of this paper is an analysis of the herding externality with forward-looking behavior. Contrary to the popular impression of incorrect herds as a market failure, we show that herding is constrained-efficient: Even when $\mathcal{DM}$s internalize the herding externality by placing very low weight on their private gain, incorrect herds obtain whenever private signals are boundedly powerful; however, they occur with a vanishing chance as the discount factor tends to one. The $\mathcal{DM}$s' lack of concern for successors affects just the extent of incomplete learning, and not its existence.

Social information is poorly aggregated by action observation,[2] as individuals may choose actions that reveal almost none of their information. But suppose that early $\mathcal{DM}$s either wished to help latecomers, or were so induced, by taking more revealing actions that better signalled their private information. *Exactly* what instructions should a planner give to the $\mathcal{DM}$'s to maximize social welfare? We provide a compact description of optimal behaviour in this constrained efficient herding model. We produce a set of indices that capture the privately estimated social value of every action. The optimal decision rule is simply to choose the action with the highest index. While they have the flavour of Gittins' (1979) multi-armed bandit indices, they also must incorporate the potential to signal to successors. So unlike Gittins' perfect information context, an action's index is not simply its present social value. Rather, individuals' signals are hidden from view, and therefore social rewards must be translated into private incentives using the marginal social value.

[1] As convergence may take infinite time (see SS), we have a *limit cascade*, as opposed to BHW's cascade.

[2] Dow (1991) and Meyer (1991) have also studied the nature of such a coarse information process for different contexts: search theory and organizational design.

We apply these indices to establish a key comparative static, that the cascade belief set strictly shrinks when $\mathcal{DM}$s grow more patient.

Our second application of the indices is to the equivalent problem of the informationally-constrained social planner, whose goal is to maximize the present discounted value of all $\mathcal{DM}$s' welfare. For the altruistic herding fiction is only worthy of study if it can be decentralized. A social planner must encourage early $\mathcal{DM}$s to sacrifice for the informational benefit of posterity. Such an outcome can be decentralized as a constrained social optimum by means of a simple set of history-dependent balanced-budget transfers. These transfers are given by our indices, and have a rather intuitive economic meaning.

This paper was originally sparked by a simple question about informational herding: Haven't we seen this before? We were piqued by its similarity to the familiar failure of complete learning in an optimal experimentation problem. Rothschild's (1974) analysis of the two-armed bandit is a classic example: An impatient monopolist optimally experiments with two possible prices each period, with fixed but uncertain purchase chances for each price. Rothschild showed that the monopolist (*i*) eventually settles down on one of the two prices, and (*ii*) selects the less profitable price with positive probability. To us, this had the clear ring of: (*i*) an action herd occurs, and (*ii*) with positive chance is misguided.

This paper also formally justifies this intuitive link. We prove that informational herding is not a new phenomenon, but a camouflaged context for an old one: myopic single person experimentation, with possible incomplete learning. Our proof respects the herding paradigm quintessence that predecessors' signals be hidden from view. In a nutshell, we replace all $\mathcal{DM}$s by agent machines that automatically map any realized private signals into action choices; the true experimenter then must furnish these automata with optimal history-contingent 'decision rules'. We therefore reinterpret actions in the herding model as the experimenter's stochastic signals, and the $\mathcal{DM}$s' decision rules as his allowed actions. We perform this formal embedding for a very general observational learning context.[3]

The organization of this paper is as follows. As we must first solve the experimentation problem anyway, it makes more sense to proceed backwards, ending with the economics. So section 2 describes a general infinite player observational learning model, and then re-interprets it as an optimal single-person experimentation model. Focusing on the finite-action informational herding model, section 3 characterizes the experimenter's limiting beliefs. The altruistic herding model is introduced in section 4, and optimal strategies are described using index rules; these are then applied for our key comparative static, as well as a description in section 5 of the optimal transfers for the equivalent planner's problem. A conclusion affords a broader perspective on our findings. Some proofs are appendicized.

[3]Such an embedding is well-known and obvious for rational expectations pricing models (eg. Scheinkman and Schechtman (1983)), since the price is publicly observed, and an inverse mapping is not required.

# 2. TWO EQUIVALENT LEARNING MODELS

In this section, we first set up a rather general observational learning model, that subsumes SS, and thus BHW and Banerjee (1992). All models in this class are then formally embedded in the experimentation framework. Afterwards, we specialize our findings.

## 2.1 Observational Learning

**Information.** An infinite sequence of decision-makers ($\mathcal{DM}$s) $n = 1, 2, \ldots$ takes actions in that exogenous order. There is uncertainty about the payoffs from these actions. The elements of the parameter space $(\Omega, \mathcal{F})$ are referred to as *states of the world.* There is a given common prior belief, the probability measure $\nu$ over $\Omega$.

The $n$th $\mathcal{DM}$ observes a partially informative private random signal $\sigma_n \in \Sigma$ about the state of the world. As shown in SS (Lemma 1), we may assume WLOG that the private signal received by a $\mathcal{DM}$ is actually his *private belief,* i.e. we let $\sigma$ be the measure over $\Omega$ which results from Bayesian updating given the prior $\nu$ and observation of the private signal. Signals thus belong to $\Sigma$, the space of probability measures over $(\Omega, \mathcal{F})$, and $\mathcal{G}$ is the associated sigma-algebra. Conditional on the state, the signals are assumed to be i.i.d. across $\mathcal{DM}$s, drawn according to the probability measure $\mu^\omega$ in state $\omega \in \Omega$. To avoid trivialities, assume that not all $\mu^\omega$ are (a.s.) identical, so that some signals are informative. Each distribution may contain atoms, but to ensure that no signal will *perfectly* reveal the state of the world, we insist that all $\mu^\omega$ be mutually absolutely continuous (a.c.), for $\omega \in \Omega$.

**Bayesian Decision-Making.** Everyone chooses from a fixed action set $A$, equipped with the sigma-algebra $\mathcal{A}$. Action $a$ earns a nonstochastic payoff $u(a, \omega)$ in state $\omega \in \Omega$, the same for all $\mathcal{DM}$s, where $u : A \times \Omega \mapsto \mathbb{R}$ is measurable. It is common knowledge that everyone is rational, i.e. seeks to maximize his expected payoff. Before deciding upon an action, everyone first observes his private signal/belief and the entire action history $h$.

Each $\mathcal{DM}$'s Bayes-optimal decision uses the observed action history and his own private belief. As in SS, we simply assume that a $\mathcal{DM}$ can compute the behaviour of all predecessors, and can use the common prior to calculate the ex ante (time-0) probability distribution over action profiles $h$ in either state. Knowing these probabilities, Bayes' rule implies a unique *public belief* $\pi = \pi(h) \in \Sigma$ for any history $h$. A final application of Bayes' rule given the private belief $\sigma$ yields the *posterior belief* $\rho \in \Sigma$.

Given the posterior belief $\rho$, the $n$th $\mathcal{DM}$ picks the action $a \in A$ which maximizes his expected payoff $\bar{u}_a(\rho) = \int_\Omega u(a, \omega) d\rho(\omega)$. We assume that such an optimal action $a = a(\rho)$ exists.[4] The solution defines an *optimal decision rule* $x$ from $\Sigma$ to $\Delta(A)$, the space of

[4]Absent a unique solution, we must take a measurable selection from the solution correspondence.

probability measures over $(A, \mathcal{A})$. Let $X$ be the space of such maps $x : \Sigma \mapsto \Delta(A)$. A rule $x$ produces an implied distribution over actions simultaneously for all private beliefs $\sigma$.

**The Stochastic Process of Beliefs.** Since the optimal $x$ depends on $\pi$, and the probability measure of signals $\sigma$ depends on the state $\omega$, the implied distribution over actions $a$ depends on both $\omega$ and the optimal decision rule $x$. The density is $\psi(a|\omega, x) \equiv \int x(\sigma)(a)\mu^\omega(d\sigma)$, and unconditional on the state, it is $\psi(a|\pi, x) \equiv \int_\Omega \psi(a|\omega, x)\pi(d\omega)$. This in turn yields a distribution over next period public beliefs $\pi_{n+1}$. Thus, $\langle \pi_n \rangle$ follows a discrete-time Markov process with state-dependent transition chances.

## 2.2 Informational Herding as Experimentation Déjà Vu

> And out of old bookes, in good faithe,
> Cometh al this new science that men lere.
> — Geoffrey Chaucer[5]

Our immediate goal is to recast the observational problem outcome as a single person optimization. A first stab brings us to the *forgetful experimenter,* who each period receives a new informative signal, takes an optimal action, and then promptly forgets his signal; the next period, he can reflect only on his action choice. But this is not a model of *Bayes-optimal* experimentation, since it assumes and in fact requires irrational behaviour. How then can an experimenter not observe the private signals, and yet take informative actions?

To give proper context to our resolution, it helps to consider McLennan (1984). This nice sequel to Rothschild's work permitted the monopolist to charge one of a continuum of prices, but assumed only two possible linear purchase chance 'demand curves'. McLennan found that the resulting uninformative price when the demand curves crossed may well eventually be chosen by an optimizing monopolist.

Rothschild's and McLennan's models give examples of *potentially confounding actions,* as introduced in EK: Easley and Kiefer (1988). In brief, such actions are optimal for *unfocused* beliefs for which they are invariants (i.e. taking the action leaves the beliefs unchanged). Of particular significance is the proof in EK (on page 1059) that with finite state and action spaces, potentially confounding actions generically do not exist, and thus complete learning must arise.[6] Rothschild and McLennan might be seen as separate anticipations of EK's general insight. Rothschild escapes it by means of a continuous state space, whereas McLennan resorts to a continuous action space. Yet there appears no escape for the herding paradigm, where both flavours of incomplete learning, limit cascades

[5] See The Assembly of Fowles. Line 22.

[6] Eg: payoffs in a one-armed bandit, with a potentially confounding safe arm, are not generic in $\mathbb{R}^2$.

| Observational Learning Model | Impatient Experimenter Model |
|---|---|
| State: $\omega \in \Omega$ | State: $\omega \in \Omega$ |
| Public Belief after $n$th $\mathcal{DM}$: $\pi_n$ | Belief after $n$ observations: $\pi_n$ |
| Optimal decision rule: $x \in X$ | Optimal action: $x \in X$ |
| Private signal of $n$th $\mathcal{DM}$: $\sigma_n$ | Randomness in the $n$th experiment: $\sigma_n$ |
| Action taken by each $\mathcal{DM}$: $a \in A$ | Observable signals: $a \in A$ |
| Density over actions: $\psi(a\|\omega, x)$ | Density over observables: $\psi(a\|\omega, x)$ |
| Payoffs: private information | Payoffs: unobserved |

Table 1: **Embedding.** This table displays how our single-type observational learning model fits into the impatient single person experimentation model.

and confounded learning (see SS), generically arise with two actions and two states. This puzzle suggests the inverse mapping that we now consider.

In recasting our general observational learning model as a single person experimentation problem, we must focus on the myopic experimenter with discount factor 0 (ruling out active experimentation). Steering away from a forgetful experimenter, we shall regard the observational learning story from a new perspective. Consider the $n$th $\mathcal{DM}$, who uses both the public belief $\pi_n$ and his private signal $\sigma_n$ in forming and acting upon his posterior beliefs $\rho_n$. We may separate these two steps by the conditional independence of $\pi_n$ and $\sigma_n$. Regard Mr. $n$ as: ($i$) observing $\pi_n$, but *not* his private signal; ($ii$) optimally determining the rule $x \in X$, and submitting it to an agent 'choice' machine; and ($iii$) letting that machine observe his private signal and take his action $a \in A$ for him. The payoff $u(a, \omega)$ is unobserved, lest that provide an additional signal of the state of the world. If private beliefs $\sigma$ have distribution $\mu^\omega$ in state $\omega$, then the experimenter chooses the same optimal decision rule $x$ described in section 2, resulting in action $a \in A$ with chance $\psi(a|\omega, x)$.

Thus, the observational learning model corresponds to a single-person experimentation model where: The state space is $\Omega$. In period $n$, the experimenter $\mathcal{EX}$ chooses an action (the rule) $x \in X$. Given this choice, a random observable statistic $a \in A$ is realized with chance $\psi(a|\omega, x)$ in state $\omega$. Finally, $\mathcal{EX}$ updates beliefs using this information alone.[7] Table 1 summarizes the embedding.

Notice how this addresses both lead puzzles. First, the experimenter never knows the private beliefs $\sigma$, and thus does not forget them. Second, incomplete learning (bad herds) are entirely consistent with EK's generic finding of complete learning for models with finite

[7]This model doesn't strictly fit into the EK mold, where stage payoffs depend only on the action and the observed signal, but (unlike here) not on the parameter $\omega \in \Omega$. This is the structure of Aghion, Bolton, Harris, and Jullien (1991) (ABHJ), who admit unobserved payoffs. Alternatively, we could posit that $\mathcal{EX}$ has fair insurance, and only sees/earns his expected payoff each period and not his random realized payoff.

action and state spaces. Simply put, actions do not map to actions but to signals when one rewrites the observational learning model as an experimentation model. The true action space for $\mathcal{EX}$ is the infinite space $X$ of decision rules.

SS considered two major modifications of the informational herding paradigm. One was to add i.i.d. noise ('crazy' preference types) to the $\mathcal{DM}$'s decision problem. Noise is easily incorporated here by adding an exogenous chance of a noisy signal (random action). SS also allowed for $T$ different types of preferences, with $\mathcal{DM}$s randomly drawn from one or the other type population. Multiple types can be addressed here by simply imagining that $\mathcal{EX}$ chooses a $T$-vector of optimal decision rules from $X^T$ with (only) the choice machine observing the task and private belief, and choosing the action $a$ as before.

# 3. THE PATIENT EXPERIMENTER

## 3.1 The Reformulated Model

From now on, we restrict ourselves to the more focused herding analytic framework — a two state, finite action setting, as in SS. More states complicates but does not enrich.

We assume a state space $\Omega = \{H, L\}$, with both states equilikely ex ante, i.e. having prior $\nu(L) = \nu(H) = 1/2$. Private belief $\sigma$ is the chance of state $H$, so that $\Sigma = [0,1]$. Let $\text{supp}(\mu)$ be the common support of each probability measure $\mu^\omega$ over private beliefs (i.e. noise for $\mathcal{EX}$'s problem). If $\text{supp}(\mu) \subseteq (0,1)$, then private beliefs are *bounded*; they are *unbounded* if $\text{co}(\text{supp}(\mu)) = [0,1]$ — i.e. if arbitrarily strong private beliefs exist. The half-bounded, half-unbounded case is a direct sum of these separate analyses.

We make the standard herding assumption of a finite action space $A = \{a_1, \ldots, a_M\}$. We assume that no action is dominated, yielding the standard interval structure that action $a_m$ is optimal exactly when the posterior $\rho$ is in some sub-interval of $[0,1]$. WLOG, we can then order the actions such that $a_m$ is myopically optimal for posteriors $\rho \in [\bar{r}_{m-1}, \bar{r}_m]$, where $0 = \bar{r}_0 < \bar{r}_1 < \ldots < \bar{r}_M = 1$.

A strategy $s_n$ for period $n$ is a map from $\Sigma$ to $X$. It prescribes the rule $x_n \in X$ which must be used, given belief $\pi_n$. The planner chooses a strategy profile $s = (s_1, s_2, \ldots)$, which in turn determines the stochastic evolution of the model — i.e. a distribution over the sequences of realized actions, payoffs, and beliefs.

**The Value Function.** Our analysis here follows ABHJ and sections 9.1–2 of Stokey and Lucas (1989). The value function $v(\cdot, \delta) : \Sigma \mapsto \mathbb{R}$ for the planning problem with discount factor $\delta$ is $v(\pi, \delta) = \sup_s E[(1-\delta)\sum_{n=1}^{\infty} \delta^{n-1} u_n | \pi]$, where the expectation is over the payoff sequences implied by $s$. Recall that $\bar{u}_m(\pi) = \pi u(a_m, H) + (1-\pi) u(a_m, L)$ denotes the expected payoff from $a_m$ at belief $\pi$. Since $\bar{u}_m$ is affine, the Bellman equation

is

$$v(\pi,\delta) = \sup_{x\in X}\left\{\sum_{a_m\in A}\psi(a_m|\pi,x)\left[(1-\delta)\bar{u}_m(q(\pi,x,a_m)) + \delta v(q(\pi,x,a_m),\delta)\right]\right\} \qquad (1)$$

where $q(\pi,x,a)$ is the Bayes-updated belief from $\pi$ when $a$ is observed and rule $x$ is applied.

A (Markov) *policy* for $\mathcal{EX}$ is a map $\xi : [0,1] \to X$ (so the rule given belief $\pi$ is $\xi(\pi)$). The optimum in a Markovian decision problem with discount factor $\delta$ exists (eg. ABHJ, Theorem 4.1), and is achieved by some such policy, generically written $\xi^\delta$. In summary:

**Lemma 1** *For any discount factor $\delta < 1$, $\mathcal{EX}$ has an optimal* policy $\xi^\delta : [0,1] \to X$.

**Interval Structure.** SS shows that the myopic experimenter maps higher beliefs into higher actions: There are thresholds $0 = \theta_0 \leq \theta_1 \leq \ldots \leq \theta_M = 1$ depending on $\pi$ alone, such that action $a_m$ is strictly optimal when $\sigma \in (\theta_{m-1},\theta_m)$, and indifference between between $a_m$ and $a_{m+1}$ prevails at the knife-edge $\sigma = \theta_m$. This is also true with patience, as Lemma 2 proves. Intuitively, not only does the interval structure respect the action order to yield high immediate payoffs, but it also ensures the greatest information value, by producing the riskiest posterior belief distribution.[8]

**Lemma 2** *For the belief $\pi$ of $\mathcal{EX}$. Any optimal rule $x \in X$ is almost surely described by thresholds $0 = \theta_0 \leq \theta_1 \leq \cdots \leq \theta_M = 1$ such that action $a_m$ is taken when $\sigma \in (\theta_{m-1},\theta_m)$, and $\mathcal{EX}$ randomizes between $a_m$ and $a_{m+1}$ when $\sigma = \theta_m$.*

*Proof:* We prove that any rule $x$ without an interval structure can be strictly improved upon. For $m_1 < m_2$, let $\Sigma_i$ $(i = 1,2)$ be those beliefs in $\Sigma$ mapped with positive probability into $a_{m_i}$. Assume to the contrary that the sets are not almost surely ordered as $\Sigma_1 \leq \Sigma_2$.

If posteriors are perversely-ordered as $q(\pi,x,a_{m_1}) > q(\pi,x,a_{m_2})$, then given our action ordering, payoffs are strictly improved with no loss of information by remapping beliefs leading to $a_{m_1}$ into $a_{m_2}$, and vice versa. That is, the myopic payoff is strictly improved, since $\bar{u}_{m_1} - \bar{u}_{m_2}$ is a decreasing function, while the continuation value is unchanged.

Next, assume that $q(\pi,x,a_{m_1}) \leq q(\pi,x,a_{m_2})$. For any $\theta \in (0,1)$, define $\tilde{\Sigma}_1(\theta) \equiv (\Sigma_1 \cup \Sigma_2) \cap [0,\theta]$ and $\tilde{\Sigma}_2(\theta) \equiv (\Sigma_1 \cup \Sigma_2) \cap [\theta,1]$. Consider then the modified rule $\tilde{x}$ which equals $x$, except that $a_{m_i}$ is taken for beliefs in $\Sigma_i(\theta)$, and where $\theta$ satisfies $\psi(a_{m_i}|\pi,x) = \psi(a_{m_i}|\pi,\tilde{x})$. (It may be necessary for $\tilde{x}$ to randomize over the two actions at belief $\theta$ to accomplish that.) Since beliefs more in favour of state $L$ are mapped into $a_{m_1}$ under $\tilde{x}$, we find $q(\pi,\tilde{x},a_{m_1}) \leq q(\pi,x,a_{m_1})$, and similarly $q(\pi,\tilde{x},a_{m_2}) \geq q(\pi,x,a_{m_2})$, with at least one inequality strict. Thus, $\tilde{x}$ yields a mean preserving spread of the updated belief versus $x$, and since the continuation value function is weakly convex, its expectation is weakly improved. But as we have just argued, the myopic payoff is strictly improved. $\square$

[8]This problem is not without history. Sobel (1953) investigated an interval structure in a simple statistical decision problem, and more recently, Dow (1991) in a two period search model.

Our earlier assumption that no action is dominated is no longer so innocent, and may in fact have some bite for very patient $\mathcal{EX}$. For the information value alone, taking dominated actions in principle provides additional signalling power. Essentially, it leaves each $\mathcal{DM}$ with a larger finite alphabet with which to signal his continuous private belief. But including such actions would invalidate the above proof, and thus perhaps the interval structure. We leave this stone unturned, acknowledging the possible loss of generality.

### 3.2 Long Run Behavior

As is generally the case with Bayesian learning, convergence is deduced by application of the martingale convergence theorem to the belief process. Not only must beliefs settle down, but also $\mathcal{EX}$ is never dead wrong about the state. A proof is found in SS.

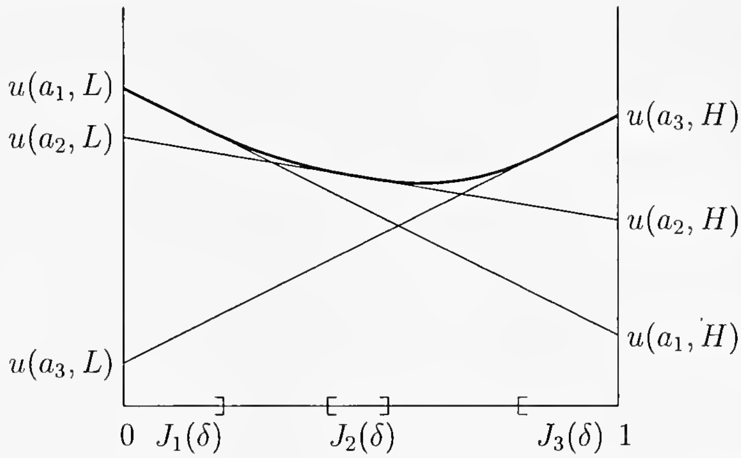
**Lemma 3** *The belief process $\langle \pi_n \rangle$ is a martingale unconditional on the state, converging a.s. to some limiting random variable $\pi_\infty$. The limit $\pi_\infty$ is concentrated on $(0,1]$ in state $H$.*

The next result is an expression of EK's Theorem 5 that the limit belief $\pi_\infty$ precludes further learning. In the informational herding model, this is only possible during a cascade, when one action chosen is chosen almost surely, and thus is uninformative. The next characterization of the stationary points of the stochastic process of beliefs $\langle \pi_n \rangle$ directly generalizes the analysis for $\delta = 0$ in SS. See figure 1 for an illustration of how the cascade sets are reflected in the shape of the optimal value function.

**Proposition 1 (Cascade Sets)** *There exist $M$ (possibly empty) subintervals of $[0,1]$, $J_1(\delta) < \cdots < J_M(\delta)$, such that $\mathcal{EX}$ optimally chooses $x$ a.s. inducing $a_m \in A$ iff $\pi \in J_m(\delta)$.*
*(a) For all $\delta \in [0,1)$, the limit belief $\pi_\infty$ is concentrated on the sets $J_1(\delta) \cup \cdots \cup J_M(\delta)$.*
*(b) With unbounded private beliefs, the extreme cascade sets are nonempty with $J_1(\delta) = \{0\}$ and $J_M(\delta) = \{1\}$, and all other $J_m(\delta)$ are empty.*
*(c) If the private beliefs are bounded, then $J_1(\delta) = [0, \underline{\pi}(\delta)]$ and $J_M(\delta) = [\bar{\pi}(\delta), 1]$, where $0 < \underline{\pi}(\delta) < \bar{\pi}(\delta) < 1$. For large enough $\delta$, all cascade sets disappear except for $J_1$ and $J_M$, while $\lim_{\delta \to 1} J_1(\delta) = \{0\}$ and $\lim_{\delta \to 1} J_M(\delta) = \{1\}$.*

*Proof:* All but the initial limit belief result are established in the appendix. To see why that one is true — that a *limit cascade* must occur, as SS call it — observe that for any belief $\hat{\pi}$ not in any cascade set, at least two signals in $A$ are realized with positive probability. By the interval structure, the highest such signal is more likely in state $H$, and the lowest more likely in state $L$. So the next period's belief differs from $\hat{\pi}$ with positive probability. Intuitively, or by the characterization result for Markov-martingale processes in appendix B of SS, $\hat{\pi}$ cannot lie in the support of $\pi_\infty$. □

Figure 1: **Typical value function.** Stylized graph of $v(\pi, \delta)$, $\delta \geq 0$, with three actions.




The proof of this result also shows that the larger is $\delta$, the weakly smaller are all cascade sets: Indeed, this drops out rather easily from the monotonicity of the value function in $\delta$. We defer asserting this result for now, and in fact Proposition 5 leverages weak monotonicity, and the index rules that we introduce later on, to deduce strict monotonicity.

**Proposition 2 (Convergence of Beliefs)** *Consider a solution of $\mathcal{EX}$'s problem.*
*(a) For unbounded private beliefs, $\pi_\infty$ is concentrated on the truth for any $\delta \in [0, 1)$.*
*(b) With bounded private beliefs, learning is incomplete for any $\delta \in [0, 1)$: Unless $\pi_0 \in J_M(\delta)$, there is positive probability in state $H$ that $\pi_\infty$ is not in $J_M(\delta)$.*
*(c) The chance of incomplete learning with bounded private beliefs vanishes as $\delta \uparrow 1$.*

*Proof:* Part $(a)$ follows from Lemma 3 and Proposition 1-a,b, and part $(b)$ just as in Theorem 1 of SS. We now extend that proof to establish the limiting result for $\delta \uparrow 1$ in part $(c)$. First, Proposition 1 assures us that for $\delta$ close enough to 1, $\pi_\infty$ places all weight in $J_1(\delta)$ and $J_M(\delta)$. The the likelihood ratio $\ell_n \equiv (1 - \pi_n)/\pi_n$ is a martingale conditional on state $H$. Because likelihood ratio $(1 - \sigma)/\sigma$ bounded above by some $\bar{\ell} < \infty$ for all private beliefs $\sigma$, the sequence $\langle \ell_n \rangle$ is bounded above by $\bar{\ell}(1 - \underline{\pi}(\delta))/\underline{\pi}(\delta)$, and the mean of $\ell_\infty$ must equal its prior mean $(1 - \pi_0)/\pi_0$. Since $\lim_{\delta \to 1} \underline{\pi}(\delta) = 0$, the weight that $\pi_\infty$ places on $J_1(\delta)$ in state $H$ must vanish as $\delta \to 1$. $\square$

Observe how incomplete learning besets even an extremely patient $\mathcal{EX}$. So this problem does not fall under the rubric of EK's Theorem 9, where it is shown that if the optimal value function $v$ is *strictly convex* in beliefs $\pi$, learning is complete for $\delta$ near 1. For here, $\mathcal{EX}$ optimally behaves myopically for very extreme beliefs: $v(\pi) = \bar{u}_1(\pi)$ for $\pi$ near 0, and $v(\pi) = \bar{u}_M(\pi)$ for $\pi$ near 1, both *affine* functions. This points to the source of the incomplete learning: lumpy signals (actions) rather than impatience. It is simply individuals' inability to properly signal their private information that frustrates the learning process.

# 4. ALTRUISTIC HERDING AND INDEX RULES

## 4.1 The Welfare Theorem

We now shift focus from the problem facing $\mathcal{EX}$ to the informational herding context with a sequence of $\mathcal{DM}$'s. For organizational simplicity, we simply first suppose that every $\mathcal{DM}$ is altruistic, but subject to the usual informational herding constraints (observable actions, unobservable signals). Define an *altruistic herding equilibrium* (AHE) as a Bayes-Nash equilibrium of the game where every $\mathcal{DM}$ $n = 1, 2, \ldots$ seeks to maximize the present discounted welfare of all posterity, themselves included: $E[(1-\delta)\sum_{k=n}^{\infty} \delta^{k-n} u_k|\pi]$. Define the expected reward of the payoff function $f$ as $E[f|\pi] \equiv \sum_{\omega\in\Omega} \pi(\omega) f(\omega)$.

The next result is quite natural, but is proved for clarity.

**Proposition 3** *For any discount factor $\delta < 1$, any optimal policy $\xi^\delta$ for $\mathcal{EX}$ is an AHE. Consequently, an AHE exists.*

*Proof:* Fix a given $\mathcal{DM}$ and a public belief $\pi$. Assume that $\xi^\delta$ is the behaviour strategy of all successors in an AHE, but that the $\mathcal{DM}$ has some rule $\hat{x}$ that is a better reply than is $\xi^\delta(\pi)$. Then $\mathcal{EX}$ can improve his value at $\pi$ by *fully* mimicking this deviation, i.e. by $(i)$ taking $\hat{x}$ in the first period and thereafter $(ii)$ continuing with $\xi^\delta$ as if the first period history had been generated by $\xi^\delta(\pi)$. This contradicts the optimality of $\mathcal{EX}$'s policy. $\square$

## 4.2 Choosing the Best Action

Recall the classical problem of the multi-armed bandit:[9] A given patient experimenter each period must choose one of $n$ actions, each having an uncertain independent reward distribution. The experimenter must therefore carefully trade-off the informational and myopic payoffs ssociated with each action. Gittins (1979) showed that optimal behavior in this model can be described by simple index rules: Attach to each action the value of the problem with just that action and the largest possible lump sum retirement reward yielding indifference. Finally, each period, just choose the action with the highest index.

We now argue that the optimal policy employed by the decision makers in an AHE has a similarly appealing form: For a given public belief $\pi$ and private belief $p$, $\mathcal{DM}$ chooses the action $a_m$ with the largest index $w_m^\delta(\pi, p)$. This measure will incorporate the social payoff, as did the Gittins index, but as privately estimated by the $\mathcal{DM}$.

Before stating the next major result, recall that $\partial g(y)$ denotes the subdifferential of the convex function $g$ at $y$ — namely, the set of all $\lambda$ that obey $g(x) \geq g(y) + \lambda \cdot (x - y)$ for all $x$. Moreover, $\partial g(y)$ is (strictly) increasing in $y$ (strict) convexity.

[9] An excellent, if brief, treatment is found in §6.5 of Bertsekas (1987).

**Proposition 4 (Index Rules)** *Fix any AHE strategy $\xi^\delta$. For $m = 1, 2 \ldots, M$, there exist $\lambda_m \in \partial v(q(\pi, \xi^\delta(\pi), a_m), \delta)$ such that the average expected present discounted value of action $a_m$ to the decision maker faced with public beliefs $\pi$ and private belief $p$ is*

$$w_m^\delta(\pi, p) = (1-\delta)\bar{u}_m(\rho(\pi, p)) + \delta \left\{ v(q(\pi, \xi^\delta(\pi), a_m), \delta) + \lambda_m \left[ \rho(\pi, p) - q(\pi, \xi^\delta(\pi), a_m) \right] \right\} \quad (2)$$

*where $\rho(\pi, p) = \pi p / [\pi p + (1-\pi)(1-p)]$ is the posterior of $\pi$ given $p$. So the optimal decision rule is to take action $a_m$ when $w_m^\delta(\pi, p) = \max_k w_k^\delta(\pi, p)$.*

*Proof:* Fix a given decision maker $\mathcal{DM}$ facing public belief $\pi$.

We now calculate the payoffs from each of the $M$ available actions $a_1, \ldots, a_M$. Action $a_m$ of the $\mathcal{DM}$ induces the public posterior belief $q_m \equiv q(\pi, \xi^\delta(\pi), a_m)$, and a corresponding state-contingent average expected discounted future payoffs of his successors, say $\bar{v}_m^L$ and $\bar{v}_m^H$. Clearly, the $\mathcal{DM}$'s expected value of any such vNM payoff stream is affine in his posterior belief $\rho$, i.e. of the form $h_m(\rho) \equiv E[v_m^\omega | \pi] = \rho \bar{v}_m^H + (1-\rho)\bar{v}_m^L$. To evaluate these payoff streams, it suffices to employ the $\mathcal{EX}$'s reckoning, since the $\mathcal{DM}$ and $\mathcal{EX}$ entertain the same future payoff objectives. Because the affine function $h_m$ presumes the behaviour which is optimal starting at belief $q_m$, we have $h_m(q_m) = v(q_m)$. Next, by employing the same strategy starting at an arbitrary public belief $r$ as at $q_m$, $\mathcal{EX}$ can achieve the value $h_m(r)$; therefore, $h_m(r) \le v(r)$. Thus, the slope of this affine function necessarily lies in the subdifferential $\partial v(q_m)$. The present value expression (2) follows. □

That $\mathcal{EX}$ can always ensure himself a payoff function tangent to the value function simply not adjusting his policy essentially was critical to this proof. This simple idea also implies convexity of the value function (eg. Lemma 2 of Fusselman and Mirman (1993)).

### 4.3 Strict Inclusion of Cascade Sets

We next use our index rule characterization to prove a key comparative static of our forward-looking informational herding model: As individuals grow more patient, the set of cascade beliefs which foreclose on learning strictly shrinks.

Before proceeding, we need two preliminary lemmata.

**Lemma 4 (Strict Value Monotonicity)** *The value function increases strictly with $\delta$ outside the cascade sets: for $\delta_2 > \delta_1$, $v(\pi, \delta_2) > v(\pi, \delta_1)$ for all $\pi \notin J_1(\delta_2) \cup \cdots \cup J_M(\delta_2)$.*

The detailed proof of this result is appendicized, but the idea is quite straightforward. Provided $\mathcal{EX}$'s strategy in some future eventuality strictly prefers a non-myopic action, his continuation value must strictly exceed his myopic value. We then show that this holds for any continuation public belief outside both cascade sets $J_m(\delta_1) \supseteq J_m(\delta_2)$. So a more patient player, who more highly weights the continuation value, will enjoy a higher value.

We also exploit the fact that we can characterize differentiability of the value function at the edge of cascade sets.[10] For context, it is in general very hard to identify primitive assumptions which guarantee the differentiability of our value function. Call rules $x$ and $y$ *equivalent* if they can be represented by the same thresholds (with associated mixing).

**Lemma 5 (Differentiability)** *Let $\hat{\pi} \neq 0, 1$ be an endpoint of cascade set $J_m(\delta)$. Assume that all rules optimal at $\hat{\pi}$ are equivalent. Then $v(\cdot, \delta)$ is differentiable in the belief at $\hat{\pi}$.*

Write the Bellman equation (1) as $v = T_\delta v$, and call $T_\delta$ the Bellman operator. As usual, $v \geq v'$ implies $T_\delta v \geq T_\delta v'$. Also, $T_\delta$ is a contraction, and $v(\cdot, \delta)$ is its unique fixed point in the space of bounded, continuous, weakly convex functions.

We can finally establish the major comparative static of this paper, that if $\mathcal{EX}$ is indifferent about foreclosing on further learning at some belief (i.e. barely in a cascade), then he strictly prefers to learn if he is slightly more patient.

**Proposition 5 (Strict Inclusion)** *Assume bounded beliefs. All non-empty cascade sets shrink strictly when $\delta$ rises: $\forall a_m \in A$, if $\delta_2 > \delta_1$ and $J_m(\delta_1) \neq \emptyset$, then $J_m(\delta_2) \subset J_m(\delta_1)$.*

*Proof:* Let $r = \inf J_m(\delta_1)$, the left edge of the cascade set. As Step 4 of the proof of Proposition 1 asserts $J_m(\delta_2) \subseteq J_m(\delta_1)$, and $J_m(\delta_1) = \{\pi | v(\pi, \delta_1) - \bar{u}_m(\pi) = 0\}$ is closed by continuity of $v(\pi, \delta_1) - \bar{u}_m(\pi)$ in $\pi$, we need only prove $r \notin J_m(\delta_2)$. There are two cases.

Case 1. Assume that at public belief $r$ and with discount factor $\delta_1$, some optimal rule $\hat{x}$ does not almost surely take action $a_m$. Instead, with positive probability, $\hat{x}$ takes some action $a_k$ producing a posterior belief $q(\pi, x, a_k)$ not in any $\delta_1$-cascade set. [For since $a_m$ is myopically optimal at $r \in J_m(\delta_1) \subseteq J_m(0)$, the optimal rule $\hat{x}$ cannot almost surely induce any other myopically suboptimal action $a_j$ ($j \neq m$) at a stationary belief.] So from Lemma 4, $v(q(\pi, \hat{x}, a_k), \delta_2) > v(q(\pi, \hat{x}, a_k), \delta_1) \geq \bar{u}_k(q(\pi, \hat{x}, a_k))$, and since we can always employ the rule $\hat{x}$ with the discount factor $\delta_2$, we must have

$$\begin{aligned} v(r, \delta_2) &\geq \sum_{a_j \in A} \psi(a_j | r, \hat{x}) \left[(1 - \delta_2)\bar{u}_j(q(r, \hat{x}, a_j)) + \delta_2 v(q(r, \hat{x}, a_j), \delta_2)\right] \\ &> \sum_{a_j \in A} \psi(a_j | r, \hat{x}) \left[(1 - \delta_1)\bar{u}_j(q(r, \hat{x}, a_j)) + \delta_1 v(q(r, \hat{x}, a_j), \delta_1)\right] = v(r, \delta_1) \end{aligned}$$

Consequently, we have $v(r, \delta_2) > v(r, \delta_1) = \bar{u}_m(r)$ and so $r \notin J_m(\delta_2)$.

Case 2. Next suppose that the optimal rule at public belief $r$ with discount factor $\delta_1$ is unique. Then the partial derivative $v_1(r, \delta_1)$ exists by Lemma 5. By the convexity of the value function, any selection from the subdifferential $\partial v(\pi)$ converges to $v_1(r, \delta_1)$

[10]We thank Rabah Amir, David Easley, Andrew McLennan, Paul Milgrom, Len Mirman, and Yaw Nyarko for private discussions about the differentiability of the value function in experimentation problems.

as $\pi$ increases to $r$. Since the optimal rule correspondence is upper hemicontinuous by the Maximum Theorem, and uniquely valued at $r$, the posterior belief $q(\pi, \xi^{\delta_1}(\pi), a_k)$ is continuous in $\pi$ at $r$ for any rule optimal selection $\xi^{\delta_1}$ and any action $a_k$.

Let $\underline{b} = \inf \operatorname{supp}(\mu)$ be the lower endpoint of the private belief distribution. As the optimal rule at $r$ almost surely prescribes action $a_m$, we let $q(r, \xi^{\delta_1}(r), a_m) = r$ and $q(r, \xi^{\delta_1}(r), a_{m-1}) = \rho(r, \underline{b})$. By their definition, $w_m^{\delta_1}(\pi, p)$ and $w_{m-1}^{\delta_1}(\pi, p)$ are then jointly continuous in $(\pi, p)$ at $(r, \underline{b})$. [In the expression for $w_{m-1}^{\delta_1}$, $\lambda_{m-1}$ lies between the slopes of $\bar{u}_1$ and $\bar{u}_M$, and is multiplied by a function that is continuous and vanishing at $(r, \underline{b})$, given $q(r, \xi^{\delta_1}(r), a_{m-1}) = \rho(r, \underline{b})$.] Also, $w_m^{\delta_1}(r, \underline{b}) \geq w_{m-1}^{\delta_1}(r, \underline{b})$ since $r$ lies in the cascade set $J_m(\delta_1)$, while $w_m^{\delta_1}(\pi, \underline{b}) < w_{m-1}^{\delta_1}(\pi, \underline{b})$ for $\pi < r$, since $r$ is the endpoint of $J_m(\delta_1)$. So $w_m^{\delta_1}(r, \underline{b}) = w_{m-1}^{\delta_1}(r, \underline{b})$ by continuity. This equality can be rewritten in a very useful form:

$$\begin{aligned} &\bar{u}_m(\rho(r, \underline{b})) - \bar{u}_{m-1}(\rho(r, \underline{b})) \\ &\quad = \delta_1 \left[\bar{u}_m(\rho(r, \underline{b})) - \bar{u}_{m-1}(\rho(r, \underline{b})) + v(\rho(r, \underline{b}), \delta_1) - v(r, \delta_1) - \lambda_m^{\delta_1}(\rho(r, \underline{b}) - r)\right] \end{aligned} \tag{3}$$

Moreover, from the previous proof of Proposition 4, $\lambda_m^{\delta_1}$ is the slope of $\bar{u}_m$, because the function $h_m(\rho) = v(r, \delta_1) + \lambda_m^{\delta_1}(\rho - r)$ evaluates the prospect of taking action $a_m$ forever.

We shall prove that $w_m^{\delta_2}(r, \underline{b}) < w_{m-1}^{\delta_2}(r, \underline{b})$, and therefore conclude that $r \notin J_m(\delta_2)$. By way of contradiction, assume that $w_m^{\delta_2}(r, \underline{b}) \geq w_{m-1}^{\delta_2}(r, \underline{b})$, i.e. $r = \inf J_m(\delta_2)$. Subtracting $w_m^{\delta_1}(r, \underline{b}) \geq w_{m-1}^{\delta_1}(r, \underline{b})$, we then have the following contradiction:

$$\begin{aligned} 0 &\geq [w_m^{\delta_2}(r, \underline{b}) - w_{m-1}^{\delta_2}(r, \underline{b})] - [w_m^{\delta_1}(r, \underline{b}) - w_{m-1}^{\delta_1}(r, \underline{b})] \\ &= (\delta_2 - \delta_1)\left[\bar{u}_m(\rho(r, \underline{b})) - \bar{u}_{m-1}(\rho(r, \underline{b})) - v(r, \delta_1)\right] \\ &\qquad + \delta_2 v(\rho(r, \underline{b}), \delta_2) - \delta_1 v(\rho(r, \underline{b}), \delta_1) - \delta_2 \lambda_m^{\delta_2}(\rho(r, \underline{b}) - r) + \delta_1 \lambda_m^{\delta_1}(\rho(r, \underline{b}) - r) \\ &> (\delta_2 - \delta_1)\left[\bar{u}_m(\rho(r, \underline{b})) - \bar{u}_{m-1}(\rho(r, \underline{b})) - v(r, \delta_1)\right] \\ &\qquad + \delta_2 v(\rho(r, \underline{b}), \delta_1) - \delta_1 v(\rho(r, \underline{b}), \delta_1) - \delta_2 \lambda_m^{\delta_1}(\rho(r, \underline{b}) - r) + \delta_1 \lambda_m^{\delta_1}(\rho(r, \underline{b}) - r) \\ &= (\delta_2 - \delta_1)\left[\bar{u}_m(\rho(r, \underline{b})) - \bar{u}_{m-1}(\rho(r, \underline{b})) - v(r, \delta_1) + v(\rho(r, \underline{b}), \delta_1) - \lambda_m^{\delta_1}(\rho(r, \underline{b}) - r)\right] \\ &= (\delta_2 - \delta_1)\left[\bar{u}_m(\rho(r, \underline{b})) - \bar{u}_{m-1}(\rho(r, \underline{b}))\right] / \delta_1 \geq 0 \end{aligned}$$

Here's a detailed justification. Under the assumption that $r \in J_m(\delta_1)$, one optimal policy $\xi_1^\delta$ induces $a_m$ almost surely at belief $r$, so that $q(r, \xi^{\delta_1}(r), a_{m-1}) = q(r, \xi^{\delta_2}(r), a_{m-1}) = \rho(r, \underline{b})$. The first equality then follows substituting (2) for each index, and using $v(r, \delta_2) = v(r, \delta_1)$ when $r \in J_m(\delta_1) \cap J_m(\delta_2)$. The second equality is simple algebra, while the final equality applies (3). The second inequality exploits $\lambda_m^{\delta_1} = \lambda_m^{\delta_2}$ (true as both are the slope of $\bar{u}_m$), and $v(\rho(r, \underline{b}), \delta_2) > v(\rho(r, \underline{b}), \delta_1)$, as given by Lemma 4. The final inequality follows since $r \in J_m(\delta_1) \subseteq J_m(0)$, so that $a_m$ is myopically optimal at $\rho(r, \underline{b})$. $\square$

# 5. CONSTRAINED EFFICIENT HERDING

Let us turn full circle, and consider once more the informational herding paradigm as played by selfish individuals. Let the realized payoff sequence be $\langle u_k \rangle$. Reinterpret $\mathcal{EX}$'s problem as that of an informationally constrained social planner $\mathcal{SP}$ trying to maximize the expected discounted average welfare $E[(1-\delta)\sum_{n=1}^{\infty}\delta^{n-1}u_n|\pi]$ of the individuals in the herding model. Observe how $\mathcal{SP}$'s and $\mathcal{EX}$'s objectives are perfectly aligned. To respect that key herding informational assumption that actions but not signals are observed, we further assume that the $\mathcal{SP}$ neither knows the state nor can observe the individuals' private signals, but can both observe and tax or subsidize any actions taken.

How does $\mathcal{SP}$ steer the choices away from the myopic solution to $\mathcal{EX}$'s problem? Given the current public belief $\pi$, if an individual takes action $a \in A$, he then receives the (possibly negative) transfer $\tau(a|\pi)$. A *constrained herding equilibrium* (CHE) is a Bayes-Nash equilibrium of the repeated game where every $\mathcal{DM}$ $n = 1, 2, \ldots$ seeks to maximize his expected one-shot myopic payoff $u(a, \pi)$ plus incurred transfers $\tau(a|\pi)$. Faced with such incentives, our proof in Lemma 2 that individuals optimally choose private belief threshold rules is still valid, for any transfers.

Since the $\mathcal{SP}$'s policy is measurable in the same observed action history as was $\mathcal{EX}$'s program, the best $\mathcal{SP}$ can do is to coax each $\mathcal{DM}$ to implement $\mathcal{EX}$'s optimal rule $x^*$. A *constrained-efficient herding equilibrium* (CEHE) is a CHE where the transfers achieve this constrained first best outcome. Existence follows at once from Lemma 1.

**Lemma 6** *For any discount factor $\delta < 1$, the optimal policy $\xi^\delta$ for $\mathcal{EX}$ is a CEHE.*

Since the private belief $\sigma$ maps into the posterior $\rho(\pi, \sigma) = \pi\sigma/[\pi\sigma + (1-\pi)(1-\sigma)]$, the selfish herder's threshold $\theta_m$ must solve the indifference equation $\bar{u}_m(\rho(\pi, \theta_m)) + \tau(a_m|\pi) = \bar{u}_{m+1}(\rho(\pi, \theta_m)) + \tau(a_{m+1}|\pi)$. So the transfer difference $\tau(a_{m-1}|\pi) - \tau(a_m|\pi)$ alone matters for how individuals trade-off between the two actions, and the $\mathcal{SP}$ can ensure that the threshold belief is optimally chosen ($\theta_m = \theta_m^*$) by suitably adjusting this net premium for taking action $a_{m-1}$ rather than $a_m$.

We want to provide some characterization of these transfers. Clearly, $\mathcal{SP}$ will not tax or subsidize actions if her desired one will be chosen anyway, i.e. for $\pi \in J_m(\delta) \subseteq J_m$. Conversely, when $\pi \notin J_m(\delta)$, $\mathcal{SP}$ perforce wishes to encourage nonmyopic actions, and some transfers intuitively will differ from zero. Indeed, $\pi \in J_m(\delta) \subset J_m$ is a strict inclusion for all $\delta > 0$ by Proposition 5, and thus transfers are not identically zero for a patient $\mathcal{SP}$.

Our action indices shed some more light on the optimal transfers, beyond the mere fact that 'experimentation' (making non-myopic choices) is rewarded. Clearly, the sum of his transfer and myopic payoffs in a CEHE must leave every individual who should be on the

knife-edge between two neighbouring optimal actions perfectly indifferent. In other words, we need $\tau(a_m|\pi) + \bar{u}_m(\rho(\pi, \theta_m)) = w_m^\delta(\pi, \theta_m)/(1-\delta)$. In fact, this condition is sufficient too, because the interval structure is optimal by Lemmata 2 and 6, and myopic incentives alone will lead inframarginal $\mathcal{DM}$'s to make the correct decisions.

**Proposition 6 (Optimal Transfers)** *For any belief $\pi$, there exist $\lambda_1 < \cdots < \lambda_M$, with $\lambda_m \in \partial v(q(\pi, \xi^\delta(\pi), a_m), \delta)$, so that the following are efficient transfers $\tau(a_m|\pi)$:*

$$\begin{aligned}\tau(a_m|\pi) &= \delta v(q(\pi, \xi^\delta(\pi), a_m)) + \delta\lambda_m[\rho(\pi, \theta_m) - q(\pi, \xi^\delta(\pi), a_m)]/(1-\delta) \\ &= w_m^\delta(\pi, \theta_m)/(1-\delta) - \bar{u}_m(\rho(\pi, \theta_m))\end{aligned}$$

Observe that incentives are unchanged if a constant is added to all $M$ transfers. Consequently, $\mathcal{SP}$ may also achieve *expected budget balance* each period: i.e. the expected contribution from everyone is zero, or $0 = \sum_{m=1}^{M} \psi(a_m|\pi, \xi^\delta(\pi))\tau(a_m|\pi)$. There is obviously a unique set of efficient transfers that satisfies budget balance.

### 5.1 Herding is Constrained Efficient

We are now positioned to reformulate the learning results of the last section at the level of actions. First a clarifying definition: We say that a *herd* arises on action $a_m$ at stage $N$ if all individuals $n = N, N+1, N+2, \ldots$ choose action $a_m$. Observe that this differs from the definition of a cascade; certainly, a cascade will imply a herd, but the converse is false. To show that herds arise, we can generalize the *Overturning Principle* of SS to this case: Claim 4 (statement and proof appendicized) establishes that for $\pi$ near $J_m(\delta)$, actions other than $a_m$ will push the updated public belief far from its current value. Thus, convergence of beliefs implies convergence of actions — or, a limit cascade implies a herd. The following is thus a corollary to Proposition 2.

**Proposition 7 (Convergence of Actions)** *In any CEHE for discount factor $\delta$:*
*(a) An ex post optimal herd eventually starts for $\delta \in [0, 1)$ and unbounded private beliefs.*
*(b) With bounded private beliefs, a herd on an action eventually starts. Unless $\pi_0 \in J_M(\delta)$, a herd arises on an action other than $a_M$ with positive chance in state $H$ for any $\delta \in [0, 1)$.*
*(c) The chance of an incorrect herd with bounded private beliefs vanishes as $\delta \uparrow 1$.*

It is no surprise that $\mathcal{SP}$ ends up with full learning with unbounded beliefs, for even selfish individuals will. More interesting is that $\mathcal{SP}$ optimally incurs the risk of an ever-lasting incorrect herd. Herding is truly a robust property of the observational learning paradigm.

# 6. CONCLUSION

This paper has shown that informational herding in models of observational learning is not such an adverse phenomenon after all: Rather, it is a constrained efficient outcome of the social planner's problem, and is robust to changing the planner's discount factor. To understand this decentralized outcome, we have then derived an expression for the social present value of each action. This formulation differs from the Gittins index because of the agency problem: Since private signals are privately observed, aligning private and social incentives entails a translation using the marginal social value. Finally, we have used this expression to prove a strict comparative static that eludes dynamic programming methods: Namely, cascade sets strictly shink as the planner grows more patient.

This paper has also discovered and explored the fact that informational herding is simply incomplete learning by a single experimenter, suitably concealed. Our mapping, recasting everything in rule space, has led us to an equivalent social planner's problem. While the *revelation principle* in mechanism design also uses such a 'rule machine', the exercise is harder for multi-period, multi-person models with uncertainty, since the planner must respect the agents' belief filtrations. While this is trivially achieved in rational expectation price settings, one must exploit the martingale property of public beliefs with observational learning, and largely invert the model. This also works for more general social learning models without action observation — provided an entire history of posterior belief signals is observed. Absent this assumption, the public belief process (howsoever defined) ceases to be a martingale, and expression as an experimentation model with perfect recall is no longer possible. This explains why our model of social learning with random sampling Smith and Sørensen (1997b) must employ entirely different techniques (Polya urns).

Of course, once informational herding is correctly understood as single-person Bayesian experimentation, it no longer seems so implausible that incorrect herds may be constrained efficient. For incomplete learning is if anything the hallmark of optimal experimentation models, even with forward-looking behaviour. This link also offers hope for reverse insights into the experimentation literature: As in EK, incomplete learning at the very least requires an optimal action $x$ for which unfocused beliefs are invariant, i.e. the distribution $\psi(a|\omega, x)$ of signals $a$ is the same for all states $\omega$. For such an invariance is clearly easier to satisfy with fewer available signals, and not surprisingly herding and all published failures of complete learning that we have seen assume a finite (vs. continuous) signal space. For instance, Rothschild (1974), McLennan (1984), and ABHJ's example are all binary signal models. More precisely, the unbounded beliefs assumption in an experimentation context corresponds to an ability to run experiments with an arbitrarily small marginal cost (eg. shifting the threshold $\theta_k$ slightly up only incurs myopic costs $o(d\theta_k)$).

# A. APPENDIX

Let the Bellman operator $T_\delta$ be given by $T_\delta v$ equals the RHS of (1). Note that for $v \geq v'$ we have $T_\delta v \geq T_\delta v'$. As is standard, $T_\delta$ is a contraction, and $v(\cdot, \delta)$ is its unique fixed point in the space of bounded, continuous, weakly convex functions.

## A.1 Proof of Proposition 1

The proposition is established in a series of steps. First, define the myopic expected utility frontier function $v_0$ by $v_0(\pi) = \max_m \bar{u}_m(\pi)$.[11]

**Step 1 (Interval Cascade Sets)** *For each $a_m \in A$, a possibly empty interval $J_m(\delta)$ exists, such that when $\pi \in J_m(\delta)$, $\mathcal{SP}$ optimally chooses $x$ with* $\operatorname{supp}(\mu) \subseteq [\theta_{m-1}, \theta_m]$, *i.e. $a_m$ occurs a.s. (learning stops). For any $\delta \in [0, 1)$, $0 \in J_1(\delta)$ and $1 \in J_M(\delta)$.*

*Proof:* For the first half, we really need only prove that $J_m(\delta)$ must be an interval. If $\pi \in J_m(\delta)$, then $a_m$ is the optimal choice, and the value is $v(\pi, \delta) = v_0(\pi) = \bar{u}_m(\pi)$. Conversely, if $v(\pi, \delta) = v_0(\pi) = \bar{u}_m(\pi)$ then $\pi \in J_m(\delta)$ and $a_m$ is the optimal choice. As $\bar{u}_m(\pi)$ is an affine function of $\pi$, and $v(\cdot, \delta)$ is weakly convex, $J_m(\delta)$ must be an interval.

For the second half, $a_1$ is myopically strictly optimal for the focused belief $\pi_n = 0$, and since it updates to $\pi_{n+1} = \pi$ a.s. no matter which rule is applied, it is also dynamically optimal for any discount factor $\delta \in [0, 1)$. A similar argument holds when $\pi_n = 1$. $\square$

**Step 2 (Iterates and Limit)** *The sequence $\{T_\delta^n v_0\}$ consists of weakly convex functions that are pointwise increasing, and converge to $v(\cdot, \delta)$. The value $v(\cdot, \delta)$ weakly exceeds $v_0$, and strictly so outside the cascade sets: $v(\pi, \delta) > v_0(\pi)$ $\forall \delta \in [0, 1)$ and $\forall \pi \notin \cup_{m=1}^M J_m(\delta)$.*

*Proof:* To maximize $\sum_{a_m \in A} \psi(a_m | \pi, x) [(1 - \delta)\bar{u}_m(q(\pi, x, a_m)) + \delta v_0(q(\pi, x, a_m))]$ over $x$ for given $\pi$, one rule $\hat{x}$ almost surely chooses the myopically optimal action. Then $q(\pi, \hat{x}, \hat{x}(\sigma)) = \pi$ a.s., resulting in value $v_0(\pi)$. Optimizing over all $x \in X$, $T_\delta v_0(\pi) \geq v_0(\pi)$ for all $\pi$. By induction, $T_\delta^n v_0 \geq T_\delta^{n-1} v_0$, yielding (as usual) a pointwise increasing sequence converging to the fixed point $v(\cdot, \delta) \geq v_0$. Finally, when $\pi$ is outside the cascade sets, by definition it is *not* optimal to almost surely induce one action. So, $v(\pi, \delta) > v_0(\pi)$. $\square$

The following either is or ought to be a folk result for optimal experimentation, but we have not found a published or cited proof of it.[12] At any rate, it is here for completeness.

**Step 3 (Weak Value Monotonicity)** *The value function is weakly increasing in $\delta$: Namely, for $\delta_1 > \delta_2$, $v(\pi, \delta_1) \geq v(\pi, \delta_2)$ for all $\pi$.*

[11] Observe how this differs from $v(\pi, 0) \equiv \sup_x \sum_m \psi(a_m | \pi, x) \bar{u}_m(q(\pi, x, a_m))$. In other words, $v(\pi, 0)$ allows the myopic individual to observe one signal $\sigma$ before obtaining the ex post value $v_0(\rho(\pi, \sigma))$.

[12] But ABHJ do assert without proof (p. 625) that the patient value function exceeds the myopic one.

*Proof:* Clearly, $\sum_{a_m \in A} \psi(a_m|\pi, x)\bar{u}_m(q(\pi, x, a_m)) \leq \sum_{a_m \in A} \psi(a_m|\pi, x)v(q(\pi, x, a_m))$ for any $x$ and any function $v \geq v_0$. If $\delta_1 > \delta_2$, then $T_{\delta_1}v_0 \geq T_{\delta_2}v_0$, since more weight is placed on the larger component of the RHS of (1). Because one possible policy under $\delta_1$ is to choose the $\xi$ optimal under $\delta_2$, we have $T^n_{\delta_1}v_0 \geq T^n_{\delta_2}v_0$. Let $n \to \infty$ and apply step 2. □

**Step 4 (Weak Inclusion)** *All cascade sets weakly shrink when $\delta$ increases: In other words, $\forall a_m \in A$, if $1 > \delta_1 > \delta_2 \geq 0$, then $J_m(\delta_1) \subseteq J_m(\delta_2)$.*

*Proof:* As seen in steps 3 and 2, $v(\pi, \delta_1) \geq v(\pi, \delta_2) \geq v_0(\pi) \geq \bar{u}_m(\pi)$ for all $\pi$, when $\delta_1 > \delta_2$. For $\pi \in J_m(\delta_1)$, we know $v(\pi, \delta_1) = \bar{u}_m(\pi)$ and thus $v(\pi, \delta_2) = \bar{u}_m(\pi)$. The optimal value can thus be obtained by inducing $a_m$ a.s., so that $\pi \in J_m(\delta_2)$. □

**Step 5 (Unbounded Beliefs)** *With unbounded private beliefs, only cascade sets for the extreme actions are empty, with $J_1(\delta) = \{0\}$ and $J_M(\delta) = \{1\}$; all other $J_m(\delta)$ are empty.*

*Proof:* SS establish for the myopic model that all $J_m(0)$ are empty, except for $J_1(0) = \{0\}$ and $J_M(0) = \{1\}$. Now apply steps 1 and 4. □

**Step 6 (Bounded Beliefs)** *If the private beliefs are bounded, then $J_1(\delta) = [0, \underline{\pi}(\delta)]$ and $J_M(\delta) = [\bar{\pi}(\delta), 1]$, where $0 < \underline{\pi}(\delta) < \bar{\pi}(\delta) < 1$.*

*Proof:* We prove that for sufficiently low beliefs it is optimal to choose a rule $x$ that almost surely induces $a_1$; the argument for large beliefs is very similar. Since action $a_1$ is optimal at belief $\pi = 0$, and is not weakly dominated, it must be the optimal choice for beliefs $\pi \leq \tilde{\pi}$, for some $\tilde{\pi} > 0$. Thus, $\bar{u}_1(\pi) = v_0(\pi)$ on $[0, \tilde{\pi}]$. Since each $\bar{u}_m$ is affine, $\bar{u}_1(\pi) > \bar{u}_m(\pi) + \underline{u}$ for all $m \neq 1$ for some $\underline{u} > 0$, and for all beliefs $\pi$ in the interval $[0, \tilde{\pi}/2]$.

No observation $a \in A$ can produce a stronger signal than any $\sigma \in \operatorname{supp}(\mu) \subseteq [\underline{\sigma}, \bar{\sigma}] \subset (0, 1)$. So any initial belief $\pi$ is updated to at most $\bar{q}(\pi) = \pi\bar{\sigma}/[\pi\bar{\sigma} + (1-\pi)(1-\bar{\sigma})]$. For $\pi$ small enough, $\bar{q}(\pi) \in [0, \tilde{\pi}/2]$ and $\bar{q}(\pi) - \pi$ is arbitrarily small, and so is $v(\bar{q}(\pi), \delta) - v(\pi, \delta)$ small by continuity of $v$ — in particular, less than $\underline{u}(1-\delta)/\delta$ for small enough $\pi$. By the Bellman equation (1), any action $a \neq a_1$ is strictly suboptimal for such small beliefs. □

**Step 7 (Limiting Patience)** *For large enough $\delta$, all cascade sets disappear except for $J_1(\delta)$ and $J_M(\delta)$, while $\lim_{\delta \to 1} J_1(\delta) = \{0\}$ and $\lim_{\delta \to 1} J_M(\delta) = \{1\}$.*

*Proof:* Fix $\delta \in [0, 1)$, and an action index $m$ $(1 < m < M)$ for which $J_m(\delta) = [\pi_1, \pi_2]$, for some $0 < \pi_1 \leq \pi_2 < 1$. Since there are informative private beliefs, $\exists \theta^* \in (1/2, 1)$ with $1 > \mu^H([\theta^*, 1]) > \mu^L([\theta^*, 1]) > 0$. We shall consider the alternative rule $x$, with interval boundaries $\theta_{m-1} = 0$, $\theta_m = \theta^*$, $\theta_{m+1} = 1$ (see Lemma 2).

Updating the prior $\pi$ with the event $\{\sigma \in [\theta^*, 1]\}$ results in the posterior belief $q(\pi) = \pi\mu^H([\theta^*, 1])/[\pi\mu^H([\theta^*, 1]) + (1-\pi)\mu^L([\theta^*, 1])]$ in state $H$. For any compact subinterval

$I \subset (0,1)$, in particular one with $I \supseteq J_m(\delta)$, there exists $\varepsilon \equiv \varepsilon(I) > 0$ with $q(\pi) - \pi \geq \varepsilon$ for all $\pi \in I$. By definition of $\varepsilon$, $q$ maps the interval $[\pi_2 - \varepsilon/2, \pi_2]$ into (but not necessarily onto) $[\pi_2 + \varepsilon/2, 1]$. Choose $\bar{u} > 0$ so large that $\bar{u}_m(\pi) < \bar{u}_{m+1}(\pi) + \bar{u}$ for all $\pi \in [0,1]$. Since $v(\pi, \delta) > \bar{u}_m(\pi)$ outside $J_m(\delta) = [\pi_1, \pi_2]$, and both are continuous in $\pi$, we may also choose $\underline{u} > 0$ so small that $v(\pi, \delta) > \bar{u}_m(\pi) + \underline{u}$ for all $\pi \in [\pi_2 + \varepsilon/2, 1]$. By step 3, we thus have $v(\pi, \delta') > \bar{u}_m(\pi) + \underline{u}$ for all $\delta' > \delta$. If $\delta' > \delta$ is so large that $(1-\delta')\bar{u} < \delta'\underline{u}$, then the Bellman equation (1) reveals that our suggested rule $x$ beats inducing $a_m$ a.s. when $\pi \in [\pi_2 - \varepsilon/2, \pi_2]$. By iterating this procedure a finite number of times, each time excising length $\varepsilon/2$ from interval $J_m(\delta)$, we see that $J_m(\delta)$ evaporates for large enough $\delta$.

If $m = 1$ or $M$, apply this procedure repeatedly: $J_m(\delta) \cap I$ vanishes for $\delta$ near 1. □

## A.2 Proof of Lemma 4

We first consider a stronger version of step 3. Call the private signal distribution *TS (Two Signals)* if its support contains only two isolated points (as is the case in BHW).

**Claim 1 (Unreachable Cascade Sets)** *Fix $\delta \geq 0$. If TS fails, then for any $\pi$ not in any $\delta$-cascade set ($\bigstar$): an action $a_m$ is taken with positive chance inducing a posterior belief $q(\pi, x, a_m)$ not in any $\delta$-cascade set. If TS holds, then ($\bigstar$) obtains for all non-cascade beliefs $\pi$ except possibly at most $M-1$ points, each the unique belief between any pair of nonempty cascade sets $J_{m-1}(0)$ and $J_m(0)$ from which both cascade sets can be reached.*

*Proof:* At a non-cascade belief $\pi$, at least two actions are taken with positive chance, and by the interval structure, some action shifts the public belief upwards while another shifts it downwards. With unbounded beliefs, $q(\pi, x, a_m)$ never lies in a cascade set; therefore, assume bounded beliefs. Let $\text{co}(\text{supp}(F)) = [\underline{b}, \bar{b}]$. Assume that $\pi$ lies between the nonempty cascade sets $J_{m'}(0)$ and $J_m(0)$, with $m' < m$, and let $\underline{\pi} = \sup J_{m'}(0)$ and $\bar{\pi} = \inf J_m(0)$. By definition of these cascade sets, $\rho(\underline{\pi}, \bar{b}) \leq \rho(\bar{\pi}, \underline{b})$. If all possible actions at $\pi$ led into a cascade set, then $\rho(\pi, \underline{b}) \leq \underline{\pi}$ and $\rho(\pi, \bar{b}) \geq \bar{\pi}$. But these inequalities can only hold with equality:

$$\rho(\rho(\pi, \bar{b}), \underline{b}) \geq \rho(\bar{\pi}, \underline{b}) \geq \rho(\underline{\pi}, \bar{b}) \geq \rho(\rho(\pi, \underline{b}), \bar{b}) = \rho(\rho(\pi, \underline{b}), \bar{b})$$

and because the outer terms coincide, as Bayes-updating commutes. So, between $J_{m'}(0)$ and $J_m(0)$ there exists at most one point $\hat{\pi}$ which can satisfy both equations; moreover, such a point exists iff $m' = m-1$ and *TS* holds. Indeed, given *TS*, we may simply choose $\hat{\pi}$ to solve $\rho(\hat{\pi}, \bar{b}) = \bar{\pi}$, while if *TS* fails, then with positive chance, a nonextreme signal is realized, and the posterior $q$ is not in a cascade set. With $\delta > 0$ we have weakly smaller cascade sets by Step 4 of the Proposition 1 proof, so a $\hat{\pi}$ failing ($\bigstar$) is even less likely to exist — in fact it would further require $\sup J_{m-1}(\delta) = \sup J_{m-1}(0)$ and $\inf J_m(\delta) = \inf J_m(0)$.

Finally, assume *TS*. Consider any $\hat{\pi}$ with reachable cascade sets $J_{m-1}(\delta)$ and $J_m(\delta)$. Then the rule $\hat{x}$ mapping $\underline{b}$ into $a_{m-1}$ (low signal to $\underline{\pi}$) and $\bar{b}$ into $a_m$ (high signal to $\bar{\pi}$) is indeed optimal. By convexity, $v(\pi,\delta)$ is at most the average of $v(\bar{\pi},\delta)$ and $v(\underline{\pi},\delta)$ (weights given by transition chances), and $\hat{x}$ achieves this average. So $v(\cdot,\delta)$ is affine on $(\bar{\pi},\underline{\pi})$. □

We now finish proving Lemma 4. From step 2 of Proposition 1's proof, $v(\pi,\delta_1) > v_0(\pi)$ for $\pi$ outside the $\delta_1$-cascade sets. Fix $\pi$ outside the $\delta_2$-cascade sets. If $\pi$ lies in a $\delta_1$-cascade set we're done, as $v(\pi,\delta_1) = v_0(\pi) < v(\pi,\delta_2)$. Suppose $\pi$ lies outside the $\delta_1$-cascade sets.

Assume first that $\pi$ satisfies (★) of Claim 1 for $\delta_1$ (and thus also for $\delta_2$). Then at least one action $a_m$ is taken with positive chance inducing a belief $q(\pi,\xi^{\delta_1}(\pi),a_m)$ not in a $\delta_1$-cascade set. Thus, $v(q(\pi,\xi^{\delta_1}(\pi),a_m),\delta_1) > v_0(q(\pi,\xi^{\delta_1}(\pi),a_m))$. Since $\delta_2 > \delta_1$,

$$v(\pi,\delta_1) = (T_{\delta_1}v(\cdot,\delta_1))(\pi) < (T_{\delta_2}v(\cdot,\delta_1))(\pi) \le (T_{\delta_2}v(\cdot,\delta_2))(\pi) = v(\pi,\delta_2) \tag{4}$$

Next assume that some $\hat{\pi}$ between $J_{m-1}(\delta_1)$ and $J_m(\delta_1)$ fails (★) for $\delta_1$. If (4) holds at $\hat{\pi}$, we are done. Assume not. Claim 1 noted that between consecutive cascade sets such $\hat{\pi}$ must be unique, and that it implied *TS*. In that case, (4) holds in a punctured neighbourhood $(\underline{\pi},\pi)\cup(\pi,\bar{\pi})$ of $\hat{\pi}$, where $\underline{\pi} = \sup J_{m-1}(\delta_1)$ and $\bar{\pi} = \inf J_m(\delta_1)$. Also, from the last paragraph of Claim 1's proof, $v(\cdot,\delta_1)$ was everywhere an affine function on $[\underline{\pi},\bar{\pi}]$, which in turn, is a supporting tangent line to the convex function $v(\cdot,\delta_2)$ at $\hat{\pi}$ (see Step 3). As it touches $v(\cdot,\delta_2)$ at $\hat{\pi}$ only, $v(\underline{\pi},\delta_2) > v(\underline{\pi},\delta_1)$ and $v(\bar{\pi},\delta_2) > v(\bar{\pi},\delta_1)$.

To find a lower bound to $v(\hat{\pi},\delta_2)$, apply rule $\hat{x}$ from Claim 1's proof at the belief $\hat{\pi}$. Since $\hat{x}$ maps $\underline{b}$ into $\underline{\pi} \in J_{m-1}(\delta_1)$ and $\bar{b}$ into $\bar{\pi} \in J_m(\delta_1)$, it yields myopic first-period values $\bar{u}_{m-1}(\underline{\pi}) = v(\underline{\pi},\delta_1)$ and $\bar{u}_m(\bar{\pi}) = v(\bar{\pi},\delta_1)$, and continuation values $v(\underline{\pi},\delta_2)$ and $v(\bar{\pi},\delta_2)$. From the right hand side of (1), this mixture is worth strictly more than $v(\hat{\pi},\delta_1)$:

$$\begin{aligned} v(\hat{\pi},\delta_1) &= \psi(a_{m-1}|\hat{\pi},\hat{x})v(\underline{\pi},\delta_1) + \psi(a_m|\hat{\pi},\hat{x})v(\bar{\pi},\delta_1) \\ &< \psi(a_{m-1}|\hat{\pi},\hat{x})\,[(1-\delta_2)v(\underline{\pi},\delta_1)+\delta_2 v(\underline{\pi},\delta_2)] + \psi(a_m|\hat{\pi},\hat{x})\,[(1-\delta_2)v(\bar{\pi},\delta_1)+\delta_2 v(\bar{\pi},\delta_2)] \end{aligned}$$

which is clearly at most $v(\hat{\pi},\delta_2)$. Given this contradiction, (4) must hold at $\hat{\pi}$. □

### A.3 Differentiability

**Continuity.** In light of the interval structure of Lemma 2, $\mathcal{SP}$ simply must determine the chances $\psi(a_m|\pi)$ with which to choose each action (i.e., what fraction of the signal space maps into each action). Thus, the choice set is WLOG the compact $M$-simplex $\Delta(A)$ — that is, the same strategy space as in our earlier general observational learning model. Since the objective function in the Bellman equation (1) is continuous in this choice vector

and in $\pi$, it follows from the Theorem of the Maximum (e.g. Theorem I.B.3 of Hildenbrand (1974)) that the non-empty optimal rule correspondence is upper hemi-continuous in $\pi$.

**Proof of Lemma 5.** Assume that $v$ is not differentiable at $\hat{\pi}$, and that all optimal rules at $\hat{\pi}$ are equivalent. We show this leads to a contradiction. In other words, if all optimal rules at $\hat{\pi}$ are equivalent, then $v$ is differentiable at $\hat{\pi}$, as asserted.

**Claim 2** $\forall \varepsilon > 0\ \exists \delta > 0$ *such that when* $|\pi - \hat{\pi}| < \delta$*, any optimal rule at* $\pi$ *induces action* $a_m$ *with probability at least* $1 - \varepsilon$ *in both states* $H, L$.

*Proof:* Since $\hat{\pi} \in J_m(\delta)$, one optimal rule at $\hat{\pi}$ induces $a_m$ with probability one. This property is shared by all rules optimal at $\hat{\pi}$. Next, if $\psi(a_m|\pi) = \pi\psi(a_m|H) + (1-\pi)\psi(a_m|L)$ is near 1, so are both $\psi(a_m|H)$ and $\psi(a_m|L)$. The claim then follows from the upper hemicontinuity of the optimal rule correspondence. $\square$

**Claim 3** $\forall N \in \mathbb{N}\ \forall \varepsilon > 0\ \exists \delta > 0$ *such that if* $|\pi - \hat{\pi}| < \delta$ *then under any optimal strategy started from* $\pi$*, action* $a_m$ *is taken for the first* $N$ *periods with probability at least* $1 - \varepsilon$ *in both states* $H, L$.

*Proof:* Fix $\eta < 1/2$. By Claim 2, for $\pi_n$ close enough to $\hat{\pi}$, action $a_m$ occurs with chance at least $1 - \eta$ in each state starting from $\pi_n$. If $a_m$ occurs, then the posterior $\pi_{n+1}$ satisfies $|\pi_{n+1} - \pi_n| \leq 4\hat{\pi}(1 - \hat{\pi})\eta$, by Bayes rule. So $|\pi_{n+1} - \pi_n|$ can be chosen arbitrarily small when $a_m$ occurs, provided $\pi_n$ is close enough to $\hat{\pi}$.

Choose the initial $\pi$ so close to $\hat{\pi}$ that if $a_m$ occurs for the next $N$ consecutive periods, the posterior belief stays close enough to $\hat{\pi}$ that $a_m$ occurs with conditional chance at least $(1 - \varepsilon)^{1/N}$ each period. In particular, we proceed as follows. Let $\rho_1 \neq \hat{\pi}$ be so close to $\hat{\pi}$ that $\rho_1(1 - \rho_1) \leq 3\hat{\pi}(1 - \hat{\pi})/2$ and at any $\pi$ within $|\rho_1 - \hat{\pi}|$ of $\hat{\pi}$, all optimal rules take $a_m$ with chance at least $(1 - \varepsilon)^{1/N}$ in each state. Let $\eta_1 = |\rho_1 - \hat{\pi}|$ and choose $\rho_2 \neq \hat{\pi}$ within $\eta_1/[8\hat{\pi}(1 - \hat{\pi})]$ of $\hat{\pi}$ and so close to $\hat{\pi}$ that $a_m$ occurs with chance at least $1 - \eta_1$ in each state from any $\pi$ within $|\rho_2 - \hat{\pi}|$ of $\hat{\pi}$. Apply this construction iteratively to choose $\eta_2 = |\rho_2 - \hat{\pi}|$ and then $\rho_3$ likewise, and then $\rho_4, \ldots, \rho_N$. If the initial belief $\pi$ lies within $|\rho_N - \hat{\pi}|$ of $\hat{\pi}$, then it stays within $|\rho_1 - \hat{\pi}|$ of $\hat{\pi}$ in the next $N$ periods provided $a_m$ occurs in each period. $\square$

We employ the machinery from the proof of Proposition 4. Any optimal strategy started at belief $\pi$ will yield some state-contingent values $\bar{v}^L$ and $\bar{v}^H$. The affine function $h(\rho)$ which has $h(0) = \bar{v}^L$ and $h(1) = \bar{v}^H$ is then a tangent to the value function at $\pi$.

Since it is optimal to take $a_m$ forever at $\hat{\pi}$, one tangent to $v$ at $\hat{\pi}$ is the affine function $h$ which intersects $u(a_m, L)$ at $\pi = 0$ and $u(a_m, H)$ at $\pi = 1$. Consider the left and right derivatives of $v$ at $\hat{\pi}$, with corresponding tangent lines $h_1(\rho)$ and $h_2(\rho)$ at belief $\rho$. One of those tangents — say, $h_1$ — must differ from $h$ (when $h_1$ differs, necessarily $m > 1$).

Define $v_1^L = h_1(0) > h(0) = u(a_m, L)$ and $v_1^H = h_1(1) < h(1) = u(a_m, H)$. Since $u(a_1, L) \geq v_1^L > u(a_m, L)$, a unique $\lambda > 0$ exists satisfying $v_1^L = \lambda u(a_1, L) + (1 - \lambda)u(a_m, L)$.

As $v$ is convex, it is differentiable almost everywhere. So let $\pi_k \uparrow \hat{\pi}$ be a sequence of beliefs converging up to $\hat{\pi}$, with the value function differentiable at each $\pi_k$. The tangent function is then uniquely determined for each $\pi_k$, and its intercepts at $\rho = 0, 1$ are the state-dependent payoffs of any optimal strategy started at $\pi_k$, namely $v^L(\pi_k) \geq v_1^L$ and $v^H(\pi_k) \leq v_1^H$. The inequalities of course follow by convexity of $v$ and $\pi_k < \hat{\pi}$.

Now choose $N$ so large and $\varepsilon$ so small that $\lambda/2 \geq 1 - (1 - \delta^N)(1 - \varepsilon)$. Note that action $a_1$ is strictly the best action in state $L$. Then by Claim 3, for all large enough $k$, the expected value $v^L(\pi_k)$ in state $L$ of the optimal strategy starting at $\pi_k$ is at most

$$\begin{aligned} v^L(\pi_k) &\leq (1 - \delta^N)(1 - \varepsilon)u(a_m, L) + [1 - (1 - \delta^N)(1 - \varepsilon)]u(a_1, L) \\ &\leq (1 - \lambda/2)u(a_m, L) + (\lambda/2)u(a_1, L) \\ &< (1 - \lambda)u(a_m, L) + \lambda u(a_1, L) = v_1^L \leq v^L(\pi_k) \end{aligned}$$

since $u(a_1, L) > u(a_m, L)$, as noted above. Contradiction. □

## A.4 Proof of Proposition 7

Near $J_m(\delta)$ we should expect to observe action $a_m$. The next lemma states that when other actions are observed they lead to a drastic revision of beliefs, *or* there was a non-negligible probability of observing some other action which would overturn the beliefs.

**Claim 4 (Overturning Principle)** *For any $\delta \in [0, 1)$, optimal $\xi^\delta$, and $J_m(\delta) \neq \emptyset$, there exists $\varepsilon > 0$ and an $\varepsilon$-neighbourhood $K \supset J_m(\delta)$, such that $\forall \pi \in K \cap (0, 1)$, either:*
*(a) $\psi(a_m|\pi, \xi^\delta(\pi)) \geq 1 - \varepsilon$, and $|q(\pi, \xi^\delta(\pi), a_k) - \pi| > \varepsilon$ for all $a_k \neq a_m$ that occur; or*
*(b) $\psi(a_m|\pi, \xi^\delta(\pi)) < 1 - \varepsilon$, and for some $a \in A$: $\psi(a|\pi, \xi^\delta(\pi)) \geq \varepsilon/M$, $|q(\pi, \xi^\delta(\pi), a) - \pi| > \varepsilon$.*

*Proof:* First, assume bounded private beliefs. By Step 6 of the proof of Proposition 1, for $\pi$ close enough to 0 or 1, the *only* optimal rule is to stop learning. Thus, we need only consider $\pi$ in some closed subinterval $I$ of $(0, 1)$. For any small enough $\eta > 0$ and $\pi$ sufficiently close to $J_m(\delta)$, we have for any $k \neq m$ : $\psi(a_k|\pi, \xi^\delta(\pi)) < 1 - \eta$. Otherwise, since the optimal rule correspondence is u.h.c., almost surely taking action $a_k$ is optimal at some $\hat{\pi} \in J_m(\delta) \subset J_m$. This is impossible, as $a_k$ incurs a strict myopic loss, and captures no information gain. Let $\text{co}(\text{supp}(F)) = [\underline{b}, \bar{b}]$. By the existence of informative beliefs, $\underline{b} < 1/2 < \bar{b}$. Let $\varepsilon > 0$ be the minimum of $\eta$, $\mu^H([\underline{b}, (2\underline{b} + 1)/4])$ and $\mu^L([(2\bar{b} + 1)/4, \bar{b}])$.

Assume $\psi(a_m|\pi, \xi^\delta(\pi)) \geq 1 - \varepsilon$ for some $\pi \in I$. Then any action $a_k \neq a_m$ is a.s. only taken for beliefs within either $[\underline{b}, (2\underline{b}+1)/4]$ or $[(2\bar{b}+1)/4, \bar{b}]$. Any such $a_k$ implies the stated overturning (selecting, if necessary, $\varepsilon$ even smaller). If instead $\psi(a_m|\pi, \xi^\delta(\pi)) < 1 - \varepsilon$, then

each action is taken with chance less than $1-\varepsilon$, and so different actions are taken at the extreme private beliefs (by the threshold structure of the optimal rule). At least one of the $M$ actions then occurs with chance at least $\varepsilon/M$ and overturns the beliefs, as claimed.

Next consider unbounded private beliefs. Assume that $\pi$ is near the cascade sets $\{0\}$ or $\{1\}$ — say $\pi$ near 0. Let the optimal policy induce with positive chance an action $a_k \neq a_1$ with $q(\pi, \xi^\delta(\pi), a_k)$ near $\pi$. Consider the altered policy that redirects private beliefs from $a_k$ into $a_1$ instead. When $\pi$ and $q(\pi, \xi^\delta(\pi), a_k)$ are near 0, this yields a boundedly positive first-period payoff gain and an arbitrarily small loss in future value (for $v$ is continuous in $q$, which shift very little, as $a_k$ was nearly uninformative). So the altered policy is a strict improvement: contradiction. Consequently, any action $a_k \neq a_1$ taken with positive chance has $|q(\pi, \xi^\delta(\pi), a_k) - \pi| > \varepsilon$ for some $\varepsilon > 0$. $\square$

For the proof of Proposition 7, we first cite the extended (conditional) Second Borel-Cantelli Lemma in Corollary 5.29 of Breiman (1968): Let $Y_1, Y_2, \ldots$ be any stochastic process, and $A_n \in \mathcal{F}(Y_1, \ldots, Y_n)$, the induced sigma-field. Then almost surely

$$\{\omega | \omega \in A_n \text{ infinitely often (i.o.)}\} = \{\omega | \sum_1^\infty P(A_{n+1}|Y_n, \ldots, Y_1) = \infty\}$$

Fix an optimal policy $\xi^\delta$. Choose $\varepsilon > 0$ to satisfy Claim 4 for all actions $a_1, a_2, \ldots, a_M$. For fixed $m$, define events $E_n = \{\pi_n \text{ is } \varepsilon\text{-close to } J_m(\delta)\}$, $F_n = \{\psi(a_m|\pi_n, \xi^\delta(\pi_n)) < 1-\varepsilon\}$, and $G_{n+1} = \{|\pi_{n+1} - \pi_n| > \varepsilon\}$. If $E_n \cap F_n$ is true, then Claim 4 scenario $(b)$ must obtain, and therefore $P(G_{n+1}|\pi_n) \geq \varepsilon/M$. So $\sum_{n=1}^\infty P(G_{n+1}|\pi_1, \ldots, \pi_n) = \infty$ on the event where $E_n \cap F_n$ occurs i.o. By the above Borel-Cantelli Lemma, $G_n$ obtains i.o. on that event almost surely. But since $\langle \pi_n \rangle$ almost surely converges by Lemma 3, $G_n$ occurs i.o. with probability zero. By implication, $E_n \cap F_n$ occurs i.o. with probability zero.

Restrict attention to the event $H$ that $\langle \pi_n \rangle$ converges to a limit in $J_m(\delta)$ and $E_n \cap F_n$ occurs only finitely many times. Then $E_n \cap G_{n+1}^c$ is eventually true on $H$, and thus so is $E_n \cap F_n^c$. But given $E_n \cap F_n^c$, all actions $a_k \neq a_m$ imply $G_{n+1}$, by the first point in Claim 4. Perforce, action $a_m$ is eventually taken on event $E_n \cap F_n^c \cap G_{n+1}^c$. Finally, sum over all $m$ to get an event of probability mass one, by Lemma 3 and Proposition 1. $\square$

## References


AGHION, P., P. BOLTON, C. HARRIS, and B. JULLIEN (1991): "Optimal Learning by Experimentation," *Review of Economic Studies*, 58(196), 621–654.

BANERJEE, A. V. (1992): "A Simple Model of Herd Behavior," *Quarterly Journal of Economics*, 107, 797–817.

Bertsekas, D. (1987): *Dynamic Programming: Deterministic and Stochastic Models.* Prentice Hall, Englewood Cliffs, N.J.

Bikhchandani, S., D. Hirshleifer, and I. Welch (1992): "A Theory of Fads, Fashion, Custom, and Cultural Change as Information Cascades," *Journal of Political Economy*, 100, 992–1026.

Breiman, L. (1968): *Probability.* Addison-Wesley, Reading, Mass.

Dow, J. (1991): "Search Decisions with Limited Memory," *Review of Economic Studies*, 58, 1–14.

Easley, D., and N. Kiefer (1988): "Controlling a Stochastic Process with Unknown Parameters," *Econometrica*, 56, 1045–1064.

Fusselman, J. M., and L. J. Mirman (1993): "Experimental Consumption for a General Class of Disturbance Densities," in *General Equilibrium, Growth, and Trade II*, ed. by J. G. et al., pp. 367–392. Academic Press, San Diego.

Gittins, J. C. (1979): "Bandit Processes and Dynamical Allocation Indices," *Journal of the Royal Statistical Society, Series B*, 14, 148–177.

Hildenbrand, W. (1974): *Core and Equilibria of a Large Economy.* Princeton University Press, Princeton.

McLennan, A. (1984): "Price Dispersion and Incomplete Learning in the Long Run," *Journal of Economic Dynamics and Control*, 7, 331–347.

Meyer, M. A. (1991): "Learning from Coarse Information: Biased Contests and Career Profiles," *Review of Economic Studies*, 58, 15–41.

Rothschild, M. (1974): "A Two-Armed Bandit Theory of Market Pricing," *Journal of Economic Theory*, 9, 185–202.

Scheinkman, J., and J. Schechtman (1983): "A Simple Competitive Model with Production and Storage," *Review of Economic Studies*, 50, 427–441.

Smith, L., and P. Sørensen (1997a): "Pathological Outcomes of Observational Learning," MIT mimeo.

——— (1997b): "Rational Social Learning Through Random Sampling," MIT mimeo.

Sobel, M. (1953): "An Essentially Complete Class of Decision Functions for Certain Standard Sequential Problems," *Annals of Mathematical Statistics*, 24, 319–337.

Stokey, N. L., and R. E. Lucas (1989): *Recursive Methods in Economic Dynamics.* Harvard University Press, Cambridge, Mass.

Date Due